PATENT COOPERATION TREATY

PCT

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

(PCT Article 36 and Rule 70)

Applicant's or agent's file reference E01/1322/WO	FOR FURTHER ACTION	See Notification of Transmittal of International Preliminary Examination Report (Form PCT/IPEA/416)
International application No. PCT/EP 03/03288	International filing date *(day/month/year)* 28.03.2003	Priority date *(day/month/year)* 28.03.2002
International Patent Classification (IPC) or both national classification and IPC G06F19/00		
Applicant EPIGENOMICS AG		

1. This international preliminary examination report has been prepared by this International Preliminary Examining Authority and is transmitted to the applicant according to Article 36.

2. This REPORT consists of a total of 6 sheets, including this cover sheet.

   ☒ This report is also accompanied by ANNEXES, i.e. sheets of the description, claims and/or drawings which have been amended and are the basis for this report and/or sheets containing rectifications made before this Authority (see Rule 70.16 and Section 607 of the Administrative Instructions under the PCT).

   These annexes consist of a total of 7 sheets.

3. This report contains indications relating to the following items:

   - I ☒ Basis of the opinion
   - II ☐ Priority
   - III ☐ Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
   - IV ☐ Lack of unity of invention
   - V ☒ Reasoned statement under Rule 66.2(a)(ii) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement
   - VI ☐ Certain documents cited
   - VII ☐ Certain defects in the international application
   - VIII ☐ Certain observations on the international application

Date of submission of the demand	Date of completion of this report
09.10.2003	29.10.2004
Name and mailing address of the international preliminary examining authority: European Patent Office - P.B. 5818 Patentlaan 2 NL-2280 HV Rijswijk - Pays Bas Tel. +31 70 340 - 2040 Tx: 31 651 epo nl Fax: +31 70 340 - 3016	Authorized Officer Türkeli, Y Telephone No. +31 70 340-2919

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

International application No. PCT/EP 03/03288

I. Basis of the report

1. With regard to the **elements** of the international application *(Replacement sheets which have been furnished to the receiving Office in response to an invitation under Article 14 are referred to in this report as "originally filed" and are not annexed to this report since they do not contain amendments (Rules 70.16 and 70.17))*:

Description, Pages

1-54 as originally filed

Claims, Numbers

1-26 received on 16.08.2004 with letter of 16.08.2004

Drawings, Sheets

1-7 as originally filed

2. With regard to the **language**, all the elements marked above were available or furnished to this Authority in the language in which the international application was filed, unless otherwise indicated under this item.

These elements were available or furnished to this Authority in the following language: , which is:

- ☐ the language of a translation furnished for the purposes of the international search (under Rule 23.1(b)).
- ☐ the language of publication of the international application (under Rule 48.3(b)).
- ☐ the language of a translation furnished for the purposes of international preliminary examination (under Rule 55.2 and/or 55.3).

3. With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, the international preliminary examination was carried out on the basis of the sequence listing:

- ☐ contained in the international application in written form.
- ☐ filed together with the international application in computer readable form.
- ☐ furnished subsequently to this Authority in written form.
- ☐ furnished subsequently to this Authority in computer readable form.
- ☐ The statement that the subsequently furnished written sequence listing does not go beyond the disclosure in the international application as filed has been furnished.
- ☐ The statement that the information recorded in computer readable form is identical to the written sequence listing has been furnished.

4. The amendments have resulted in the cancellation of:

- ☐ the description, pages:
- ☐ the claims, Nos.:
- ☐ the drawings, sheets:

5. ☐ This report has been established as if (some of) the amendments had not been made, since they have been considered to go beyond the disclosure as filed (Rule 70.2(c)).

   (Any replacement sheet containing such amendments must be referred to under item 1 and annexed to this report.)

6. Additional observations, if necessary:

V. Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

1. Statement

Novelty (N)	Yes:	Claims	9-19, 25, 26
	No:	Claims	1-8, 20-24
Inventive step (IS)	Yes:	Claims	none
	No:	Claims	1-26
Industrial applicability (IA)	Yes:	Claims	1-26
	No:	Claims	none

2. Citations and explanations

   see separate sheet

Reference is made to the following documents:

D2: 'Biological Fluid Analysis Using Distance Outlier Detection', WO 97 06418 A (BOEHRINGER MANNHEIM CORP) 20 February 1997 (1997-02-20)

D4: HUBERT ET AL, 'A fast method for robust principal components with applications to chemometrics', Chemometrics and Intelligent Laboratory Systems, 2002, 60, 101-111.

1. The document D4 is for the first time mentioned in this International Preliminary Examination Report.

2. The present application does not meet the criteria of Article 33(1) PCT, because the subject-matter of claims 1 and 20 is not new in the sense of Article 33(2) PCT.

2.1 The document **D2** is regarded as being the closest prior art to the subject-matter of claim 1, and discloses (the references in parentheses applying to this document):

A method of verifying and controlling assays for the analysis of biological fluid analytes ("collecting high quality data", "removing outliers resulting from statistical or systematic errors", page 2, par. 1; "to ensure accurate and consistent results", page 4, par. 2)

by means of statistical process control (using Hotelling's T statistics to identify degradation/outliers anticipates a statistical process control and calibration data set corresponds to the historical data set; page 9, par. 2) comprising the following steps:

a) defining a reference data set ("calibration data", page 26, par. 1)

b) defining a test data set ("unknown samples", page 26, par. 1)

and reducing the data dimensionality of the reference and test data set by means of robust embedding of the values into a lower dimensional representation (page 7, par 4; "selecting optimal number of factors with robust predictive abilities", page 23. par. 2; Figure 4)

c) determining the statistical distance between the reference data set and test data set or elements or subsets thereof (Hotelling's T statistics, Mahalanobis distance or Robust distance, page 8, par. 3 - page 9, par.2; page 26, par. 1)

d) identifying individual elements or subelements of the test dataset which have a statistical distance larger than that of a predetermined value (page 9, par.2; page 26, par. 1; claim 11).

2.2 It is to be noted that D2 does not explicitly refer to "nucleic acids"; instead, it discloses "biological fluid analytes". Such an expression does not exclude the possibility of controlling nucleic acid variations since analytes are substances being measured in an analytical procedure. Moreover, the feature of representing the elements of the reference and test sets by nucleic acid measurements does not relate to the method of claim 1, since monitoring the test data by measuring deviations from the reference data is independent of the nature of variables.

2.3 It is also to be noted that merely using the word "robust" does not signify any features providing modifications to a method of dimension reduction. Since no additional specific limitations are implied, the method of D2 employing principal component analysis which is a sound and robust method for capturing the direction of variations in the data for dimension reduction anticipates claim 1.

2.4 Even if the expression "robust principle component analysis" could be taken in the sense of comprising the specific method steps of robust PCA (claim 9) the following should be observed:

Claim 1 would define the use of "robust principle component analysis" for performing the step of dimensionality reduction. The method of D2 performs classical principal component analysis where the eigenvalues obtained from the covariance matrix are sorted and those accounting for most of the variation in the data are selected. However, it is generally known that when the data contains outliers, the covariance matrix is impaired and the eigenvalues might not describe the data correctly. When trying to implement the method of D2, the person skilled in the art is confronted with this problem since D2 discloses the presence of outliers. In order to overcome this problem, the skilled person would look for a method to improve dimensionality reduction in the presence of outliers in the relevant field.

There, the skilled person will come across **D4**, which discloses using robust principle component analysis for reduction of dimensionality in case the data involves anomalous observations.

The skilled person will incorporate the teaching of D4 in the method of D2 which

already discloses constructing a robust model to detect the outliers using PCA and employ robust PCA for dimensionality reduction without exercising any inventive skills.

As a result, even with such hypothetical interpretation of claim 1, its subject-matter would not involve an inventive step in the sense of Article 33(3) PCT.

2.5 The subject-matter of claim 20 is the same as that of claim 1 in the form of steps of a computer program corresponding to the steps of the method. Therefore, claim 20 is also not new over the disclosure of D2 in the sense of Article 33(2) PCT.

3. Dependent claims 2-19 and 21-26 do not contain any features which, in combination with the features of any claim to which they refer, meet the requirements of the PCT in respect of novelty and/or inventive step.

4. The claims are susceptible of industrial application in the sense of Article 33(4) PCT since they relate to the technical field of bioinformatics.

I/we claim:

~~1. A method of verifying and controlling assays for the analysis of nucleic acid variations by means of statistical process control, characterized in that variables of each experiment are monitored by measuring deviations of said variables from a reference data set and wherein said experiments or batches thereof are indicated as unsuitable for further interpretation if they exceed predetermined limits.~~ new claim 1

2. A method according to claim 1 when said nucleic acid variations are cytosine methylation variations.

3. A method according to claims 1 and 2 wherein said statistical process control is taken from the group comprising multivariate statistical process control and univariate statistical process control.

~~4. A method according to claims 1 to 3 comprising the following steps~~
~~a) defining a reference data set~~
~~b) defining a test data set~~
~~c) determining the statistical distance between the reference data set and test data set or elements or subsets thereof~~
~~d) identifying individual elements or subsets of the test dataset which have a statistical distance larger than that of a predetermined value.~~

~~5. The method according to claim 4, further comprising in step b)~~
~~reducing the data dimensionality of the reference and test data set by means of robust embedding of the values into a lower dimensional representation.~~

1. A method of verifying and controlling assays for the analysis of nucleic acid variations by means of statistical process control comprising the following steps

a) defining a reference data set

b) defining a test data set and reducing the data dimensionality of the reference and test data set by means of robust embedding of the values into a lower dimensional representation

c) determining the statistical distance between the reference data set and test data set or elements or subsets thereof

d) identifying individual elements or subelements of the test dataset which have a statistical distance larger than that of a predetermined value.

4. ~~6.~~ The method according to claim ~~5~~ 1 wherein step b) is carried out by calculating the embedding space using one or both of the reference and the test data sets.

5. ~~7.~~ The method according to one of claims ~~4~~ 1 to ~~6~~ 4 further comprising,

e) further investigating said identified elements or subsets of the test dataset to determine the contribution of individual variables to the determined statistical distance.

6. ~~8.~~ The method according to one of claims ~~4~~ 1 to ~~7~~ 5 further comprising,

e) excluding said identified experiments or batches thereof from further analysis.

7. ~~9.~~ The method of claim ~~4~~ 1 wherein in step d) said statistical distance is calculated by means of one or more methods taken from the group consisting the Hotelling's T^2 distance between a single test measurement vector and the reference data set, the Hotelling'-T^2 distance between a subset of the test data set and the reference data set, the distance between the covariance matrices of a subset of the test data set and the covariance matrix of the reference set, percentiles of the empirical distribution of the reference data set and percentiles of a kernel density estimate of the distribution of the reference data set, distance from the hyperplane of a nu-SVM, estimating the support of the distribution of the reference data set.

8. ~~10.~~ The method according to one of claims ~~claim 5 and 6~~ 1 to 7 wherein the data dimensionality reduction is carried out by means of principle component analysis.

9. ~~11.~~ The method according to one of claims ~~claim 5, 6 and 10~~ 1 to 7 wherein the data dimensionality reduction step comprises the following steps

i) Projecting the data set by means of robust principle component analysis

ii) Removing outliers from the data set according to their statistical distances calculated by means of one or more methods taken from the group consisting

of: Hotelling's T^2 distance; percentiles of the empirical distribution of the reference data set; Percentiles of a kernel density estimate of the distribution of the reference data set and distance from the hyperplane of a nu-SVM, estimating the support of the distribution of the reference data set.

iii) Calculating the embedding projection by standard principle component analysis and projecting the cleared or the complete data set onto this basis vector system.

10. ~~12.~~ The method according to one of claims ~~4~~ 1 to ~~11~~ 9 wherein at least one of the variables measured in steps a) and b) is determined according to the methylation state of the nucleic acids.

11. ~~13.~~ The method according to one of claims ~~4~~ 1 to ~~11~~ 9 wherein at least one of the variables measured in step a) and b) is determined by the environment used to conduct the assay.

12. ~~14.~~ The method according to one of claims ~~4~~ 1 to ~~11~~ 9 wherein said data sets comprises one or more variables selected from the group comprising mean background/baseline values; scatter of the background/baseline values; scatter of the foreground values, geometrical properties of the array, percentiles of background values of each spot and positive and negative assay control measures.

13. ~~15.~~ A method according to one of claims ~~4~~ 1 to ~~14~~ 12 wherein the reference data set is the complete series of experiments being analysed. ~~[(make it explicit in the description that the test set can be a subset of the reference data set.)]~~

14. ~~16.~~ A method according to one of claims ~~4~~ 1 to ~~14~~ 12 wherein the reference data set is derived from experiments carried out separately to those of the test data set.

15. ~~17.~~ A method according to one of claims ~~1~~ 1 to ~~14~~ 12 wherein the reference data set is derived from a set of experiments wherein the value of each variable of each experiment is either within a predetermined limit or optimally controlled.

16. ~~18.~~ A method according to one of claims ~~1~~ 1 to ~~17~~ 15 further comprising the generation of a document comprising said elements or subsets of the test data determined according to step d) of claim 4.

17. ~~19.~~ A method according to claim ~~18~~ 16 wherein said document further comprises the contribution of individual variables to the determined statistical distance.

18. ~~20.~~ A method according to claims ~~18~~ 16 and ~~19~~ 17 wherein said document is stored on a computer readable format.

19. ~~21.~~ A method according to one of claims 1 to ~~20~~ 18 wherein said method is implemented by means of a computer.

20. ~~22.~~ A computer program product for the verifying and controlling assays for the analysis of nucleic acid variations comprising

a) a computer code that receives as input a reference data set

b) a computer code that receives as input a test data set

c) a computer code that determines the statistical distance between the reference data set and test data set or elements or subsets thereof

d) a computer code that identifies individual elements or subsets of the test dataset which have a statistical distance larger than that of a predetermined value

e) a computer readable medium that stores the computer code.

f) a computer code that reduces the data dimensionality of the reference and test data set by means of robust embedding of the values into a lo dimensional repres

~~23. The computer program product of claim 22 further comprising~~

~~f) a computer code that reduces the data dimensionality of the reference and test data set by means of robust embedding of the values into a lower~~

~~dimensional representation.~~

~~21.~~ 24. The computer program product of claim ~~22~~ 20 characterised in that the embedding space is calculated using one or both of the reference and the test data sets.

~~22.~~ 25. The computer program product of claims ~~22 to 24~~ 20 and 21 further comprising,

g) a computer code that investigates said identified elements or subsets of the test dataset to determine the contribution of individual variables to the determined statistical distance.

~~23.~~ 26. The computer program product of claims ~~22~~ 20 to ~~25~~ 22 wherein said statistical distance is calculated by means of one or more methods taken from the group consisting the Hotelling's T^2 distance between a single test measurement vector and the reference data set, the Hotelling'-T^2 distance between a subset of the test data set and the reference data set, the distance between the covariance matrices of a subset of the test data set and the covariance matrix of the reference set, percentiles of the empirical distribution of the reference data set and percentiles of a kernel density estimate of the distribution of the reference data set, distance from the hyperplane of a nu-SVM, estimating the support of the distribution of the reference data set.

~~24.~~ 27. The computer program product of claims ~~23 and 24~~ 20 to 23 wherein the data dimensionality reduction is carried out by means of principle component analysis.

~~25.~~ 28. The computer program product of claims ~~23, 24 and 27~~ 20 to 24 wherein the data dimensionality reduction step comprises the following steps

i) Projecting the data set by means of robust principle component analysis

ii) Removing outliers from the data set according to their statistical distances calculated by means of one or more methods taken from the group consisting of: Hotelling's T^2 distance; percentiles of the empirical distribution of the

reference data set; Percentiles of a kernel density estimate of the distribution of the reference data set and distance from the hyperplane of a nu-SVM, estimating the support of the distribution of the reference data set.

iii) Calculating the embedding projection by standard principle component analysis and projecting the cleared or the complete data set onto this basis vector system.

26. ~~29~~. The computer program product of claims ~~22~~ 20 to ~~28~~ 25 further comprising a computer code that generates a document comprising said elements or subsets of the test data determined according to ~~step~~ d) of claim ~~22~~ 20.